Modernity and morality in Habermas's discourse ethics

Article (Accepted Version)

Modernity and Morality in Habermas's Discourse Ethics [†]

I. Introduction

One of the features that marks out Habermas's Discourse Ethics from most other contemporary moral theories is the extent to which it is informed by social scientific research in cognate areas of sociology, anthropology, and psychology. This has meant that from its inception Habermas's conception of *morality* has been hand and glove with a conception of *modernity* and with a theory of *modernization.* The moral theory forms part of a wider social theory. I take it that this is a strength, not just a peculiarity of Discourse Ethics. For much of moral philosophy after Kant, despite Hegelian protestations, has been guilty of neglecting the historical, social and cultural dimension of the phenomenon of moral normativity it explicates.

As the programme of Discourse Ethics has developed since the early 1980s so the constellation of moral theory and modernization theory has altered. Originally Discourse Ethics is conceived as a programme of *philosophical* justification of the moral principle or the moral standpoint (**MCCA,** pp. 43, 78-86, 96).[1] The formal derivation of principle (U) from non-moral premises is central to this programme. If the formal derivation goes through, then (U) can be justified on the non-moral grounds of Habermas's theory of communicative action and the pragmatic theory of meaning.[2] Thus, according to the original programme of Discourse Ethics a normative moral theory falls out of a pragmatic theory of the meaning of utterances. One of my aims in this paper is to show how and why the promised formal derivation of (U) from non-moral premises fails.

As far as Discourse Ethics is concerned this is an important and unresolved issue in its own right. But I also want to elaborate the wider significance of this failure.

The point I have in mind is this: as originally conceived, the central principle of Discourse Ethics, principle (U), is justified independently of Habermas's theory of modernity.[3] The theory of modernization, as I outline in II below, provides historical and genealogical arguments which show how morality developed from the substantive value-laden tradition into a procedural conception of rightness as impartial justification. It narrates the historical and cultural genesis of a conception of moral rightness, but does not justify that conception. At most, modernization theory, if true, lends indirect corroborative support to Discourse Ethics that is both weak and, importantly, surplus to requirements.[4] Communication theory and the pragmatic theory of meaning alone justify the moral principle and explicate the meaning of moral rightness.

According to the original programme then, modernization theory, if true, provides weak indirect evidence that the conception of morality offered by Discourse Ethics is correct. However, the conception of morality offered by Discourse Ethics, in so far as it is independently justified, lends strong justificatory support to modernization theory. The former, narrower theory illustrates the central thesis of the latter, more general theory. The conception of morality offered by the original programme of Discourse Ethics bears out the generalization contained in what I call *The Modernity Thesis*. This thesis, one that reverberates throughout Habermas's social theory, is that:

> Modernity can and will no longer borrow the criteria by which it takes its orientation from the models supplied by another epoch: *it has to create its normativity out of itself*. (**PDM**, p.7)

The original programme of Discourse Ethics offers a picture of morality in which the sources of moral normativity are contained in the formal pragmatic preconditions of speech oriented to reaching consensus, and are not drawn from a substantive conception of the good life internal to a particular tradition or form of ethical life.

This outline shows that Habermas's proposed formal derivation of (U) from non-moral premises does two things: it shapes the original programme of Discourse Ethics and it provides a justificatory support for modernization theory. The trouble is that neither Habermas nor any of his followers have so far managed to provide the formally valid derivation of (U) promised by the original programme. Habermas is aware of this lacuna and has recently proposed a weaker justification of principle (U), by abduction, *in lieu* of its formal-logical deduction (**DEA,** p.60).[5]

In III below I show why the proposed formal derivation of (U) cannot be provided. In IV I argue that the weaker abduction of (U) rests much more firmly on considerations of modernization theory than Habermas is prepared to admit. Finally, in V, I sketch the upshot of this alteration for the original programme of Discourse Ethics, namely that Habermas must abandon the forlorn task of convincing the moral sceptic and sticks the more feasible one of explicating and confirming the self-understanding of moral agents. In turn this alters the relation of the theory of morality to the theory of modernity: The Discourse Theory of Morality is no longer justified independently of modernization theory. Consequently the theory of morality no longer offers justificatory support to modernization theory: it is at least partly justified by modernization theory.

II. Modernization Theory and Programme of Discourse Ethics

(1) I shall begin by outlining Habermas's theory of modernization and its relation to his moral theory. Habermas develops his concept of modernity through a critical engagement with the concept of rationalization in Hegel, Marx, Durkheim, Weber, Lukacs, Horkheimer and Adorno. He narrates a general and plausible story concerning the emergence of secular morality as the historical consequence of a monotheistic Judaeo-Christian tradition whose values and norms presuppose the existence of an objectively good and just way of life. That way of life is recommended by a God who is both the omnipotent creator of an ordered cosmos and the absolutely just and good omniscient saviour of human kind. In this tradition each human being has a dual role, as a member of religious community of neighbours, and as an individual whose salvation depends on God's judgment. This duality is reflected in two aspects of morality: (i) universal respect for others (and accountability to all others) and (ii) the absoluteness or unconditionality of moral requirements.

According to Habermas two world-historical shifts inaugurate the transition to a modern conception of morality. First, the shattering of this religious tradition and the pluralization of conceptions of value under conditions of multiculturalism result in the separation of the notion of justice from a particular concrete conception of the good - the *ethos* of the Christian community. Second, the demise of the metaphysical conception of essence and the gradual transferal of epistemic authority to the natural sciences fundamentally alter the meaning of morality. The core conception of morality preserves itself under modern conditions by harnessing the idealized procedure of discourse as a formal standard of impartial justification that any valid substantial norm must

meet. Thus moral discourse replaces the example of a 'missing "transcendent good"' (**DEA**, p.58).

The story Habermas relates does not describe the singular fate of the Western Occidental tradition, but a more general process of the detachment of forms of normative authority from religious world-views.[6] Habermas pays particular attention to two further tendencies that I have so far omitted to mention: the rise of individual autonomy or what Hegel calls 'subjective freedom',[7]and the differentiation of spheres of rationality, the increasing 'autonomy' of aesthetic, ethical-legal and scientific rationality.[8]

The 'modernity' that results from this process presents an ambiguous legacy for modern individuals. On the positive side, their sphere of freedom is greatly increased. The power of the state, once uncoupled from religion and tradition, is held in check by publicly accessible criteria of legitimation: e.g. whether or not its affairs are run in the interests of those who comprise it. Thus modernity presents an opportunity for modern subjects to renew patterns of meaning and social interaction on a basis that promises stability, transparency and accountability.[9] On the negative side, these increases in subjective freedom and in the accountability of suprasubjective structures of authority are bought at a high price: the social deracination of individual subjects and their increasing vulnerability to the disciplinary effects impersonal systems of administration and to the vagaries of an ever more powerful capitalist economy.[10] Habermas diagnoses the negative outcome of modernization, the *social pathologies* of modernity, in the extent to which systems of 'instrumental action' corrode the repository of 'communicative action' in the life-world which is the basis of cultural reproduction, socialization and social integration, and thus sever at the

root the opportunities that modernity presents (**TKH2**, p.449-548/**TCA2**, p. 303-374).

Habermas's new approach allows him to correct the one-sided, negative analysis of rationalization which runs through Weber to the Frankfurt School. Habermas's analysis is no longer focused exclusively on the subject *qua* victim (and also in a sense *qua* perpetrator) of the ravages of social rationalization. For Habermas modernity is analysed in terms of the relation between the autonomous systems of money and power - as the embodiments of 'instrumental rationality' - and the life-world - as the embodiment of 'communicative rationality'. This approach makes salient the degree to which discourse - in particular *moral* discourse - is able to compensate for the demise of religious traditions as a common source of meaning, value, and belief by replenishing the basis of meaning. Thereby discourse discharges the tasks of socialization and social integration, and eases the burden of legitimation that now falls on all aspects of modern forms of life (**PNK**, p.226). This, I take it, is the crucial positive implication of Habermas's *Modernity Thesis*.[11]

In Habermas's theory of modernization, the tasks of the stabilization and integration of society fall primarily to practical (i.e. moral) discourse.[12] Under modern conditions 'processes of social integration are increasingly decoupled from apparently natural tradition' whilst on the institutional level 'universal moral principles and procedures of law-making replace traditional values and norms' (**PNK**, p. 226). Moral discourse can do this because, as a practice that is oriented to reaching consensus, it exercises a legitimating function and because, under modern conditions, mass loyalty and social integrity follow legitimacy.

(2) Habermas's thesis that modernity cannot borrow its sources of normativity from the past, but henceforth '*has to create its normativity out of itself*' (**PDM**, p. 7) captures the central idea of his modernization theory. The thesis is familiar to anyone who knows Habermas's work. It forms the basis of the phenomenological analysis of the time-consciousness that Habermas takes to be characteristic of modernity (**PDM**, pp. 1-44). It is central to his historical analysis of the origins of the modern epoch.[13] And it forms the internal goal towards which the developmental logic of of normative social structures unfolds: in post-traditional or post-conventional societies individual agents judge the validity of established, norms, rules or practices according to self-chosen principles.

> The liberated subjects, no longer bound and directed by traditional roles, have to create binding obligations by dint of their own communicative efforts. (**PNK**, p.231)[14]

There is no denying the centrality of *the modernity thesis* in Habermas's work. But what is the upshot of the thesis for the Discourse Theory of Morality? What makes a conception of morality distinctively modern?

Habermas's conception of morality illustrates the *Modernity Thesis* in roughly the following way. Habermas contends that the historical shift from a traditional or pre-modern to a post-traditional or modern conception of morality is accompanied by a by fundamental shift from a realist moral and evaluative world-view (Habermas tends to call it a 'metaphysical' world-view) to an idealist (or 'post-metaphysical') moral world view. Before the onset of modernity agents supposedly act under the assumption that there is a single objective moral good for man, to which true moral utterances correspond and false moral utterance do not. This assumption, according to Habermas, is an illusion, for moral goodness

is not an objective part of the fabric of the world. However the illusion that it is - that it exists independently of moral agents - is effective in coordinating action. Thus the ideal, (non-objective) nature of moral goodness is masked by the existence of culturally homogeneous communities with a shared set of 'thick' moral concepts and value conceptions which effectively prevent moral agents from discovering that the moral world depends on their attitudes. However, with the advent of modernity, the illusion of the objectivity of moral goodness is unmasked.

According to Habermas, Kant takes the decisive first step towards a modern conception of morality. Kant's first formulation of the categorical imperative locates the source of normativity not in the substantive values embodied in concrete maxims of action but in the criterion of universalizability in virtue of which those maxims are incorporated into the will. Kant's ethics make clear that the legitimacy of moral norms derives from their rational structure not their substantive content. We can call this first step *proceduralism*.

However, according to Habermas, Kant mistakenly assumes that the procedure by which moral norms are selected takes place somehow inside each solitary individual. He is blind to the intersubjective or social nature of reason. Unlike Kant Discourse Ethics locates the rational standards by which moral norms are tested for their validity in the conditions under which speakers and actors can reach intersubjective agreement in discourse, about the meaning of their moral utterances. On this point Habermas cites Thomas McCarthy's concise reformulation of his own Kantian position:

> Rather than ascribing as valid to all others any maxim that I can will to be a universal law, I must submit my maxim to all others for the purposes of discursively testing its claim to universality. The emphasis shifts from what

> each can will without contradiction to be a general law, to what all can will in agreement to be a universal norm. (**MCCA**, p.67 & **HCD**, p.257)

We can call this second step *intersubjectivism*. Thus in Habermas's view, as a result of modernization the validity of moral norms comes to be seen as a procedural, *intersubjectively ideal* construction, rather than as a real, objective property of actions. Moral utterances do not correspond to a mind-dependent reality, nor need they:

> For discourse can, thanks to the normative content of its communicative presuppositions, *create out of itself* the constraints, which are imposed on the practice of justification by the projection of a moral universe. (**RW**, p.205 my translation)[15]

Under modern conditions moral agents come to know that the amenability to 'rationally motivated consensus' in discourse is not just good evidence for, it is *constitutive* of, the validity of moral norms.

One reason why it is important that Habermas illustrate the *modernity thesis* with the example of morality is that the modernity thesis is itself a generalization of an originally aesthetic concept. For he is aware that the phenomenon of 'modernity' first emerges historically in the Eighteenth Century with 'the process of the detachment from the models of ancient art' (**PDM**, p.8). If, as Habermas claims, the process of modernization is a general one, not particular to the history of Occidental Rationalism, then the same process should manifest itself in each of the different spheres of value that modernity spawns. The example of morality shows that the phenomenon is not confined to the aesthetic sphere. It is thus good evidence for the existence of a general phenomenon that can be the proper object of the theory of modernization. Howevr it is good evidence only providing that the conception of morality in play is justified independently of the theory of modernization. Otherwise the example of morality presupposes what it is supposed to illustrate - a certain conception of modernity.

III. Principle (U) and The Programme of Discourse Ethics

Now we can turn to the question of the derivation of the moral principle. The Programme of Discourse Ethics as originally conceived is a programme of the *philosophical* or 'moral-theoretical justification of the moral point of view' (**OCCM**, p.347: **DEA**, p.59) or 'of the moral principle' (**MCCA**, p.78). The central aim of Discourse Ethics is to *justify* the moral principle (U). A recent formulation of (U) states that:

> a norm is valid if and only if the foreseeable consequences and side effects of its general observance for the interests and value-orientations of *each*

individual could be freely accepted *jointly* by *all* concerned. (**OCCM**, p. 354/ **DEA**, p.60)

(U) is a rule of argumentation that that makes agreement possible. According to Habermas's original (1983) conception of Discourse Ethics (U) is sufficiently justified if it can be derived from the following two premises:

(1) the normative, (but non-moral) preconditions of argumentation in general (**MCCA** 92)[16]

(2) a 'weak (i.e. non-moral GF) idea of normative justification' or 'the conception of normative justification in general as expressed in (D).' (**MCCA** 92, 97, 198)[17]

Note that the broader question of the *justification* of (U) depends on, but does not consist in, the narrower question of the formal-logical derivation of (U) from just these premises. (The logical derivation of (U) is a sufficient but not a necessary condition of its justification.) In his seminal 1983 essay 'Notes on a Programme of Philosophical Justification' (**MCCA** 43-116) Habermas claims that 'the programmatic justification of Discourse Ethics requires all of the following:

1. A definition of a universalization principle that functions as a rule of argumentation.

2. The identification of pragmatic presuppositions of argumentation that are inescapable and have a normative content.

3. The explicit statement of that normative content (e.g. in the form of discourse rules)

4. Proof that a relation of material implication holds between steps (3) and (1) in connection with the idea of the justification of norms.' (**MCCA** 97)

This justification programme of the moral standpoint presupposes a formal-logical derivation of (U) from the two above-mentioned premises. A 'material implication', as mentioned in step 4 above, is what is represented by the truth-functional connective, '⊃' or '→', meaning roughly, 'if, then'. Strictly speaking nothing is 'derived' merely by standing on the right-hand side of that symbol.[18] What Habermas means by step 4 is that there is a formally valid inference to (U) from premises (1) and (2). So if premises (1) - the necessary pragmatic preconditions or rules of discourse - and (2) - 'the conception of normative justification in general as expressed in (D)' - are true, then (U) can be derived by *modus ponens* in the following way.[19]

(1)R (rules of discourse)

(2)D (discursive conception of justification in general)

(3)(R and D)

(3*) if (R and D), then (U)

---

(4) (U)

This argument is no doubt formally valid. But we need more than formal validity to establish the truth of the conclusion. To do that we have to establish that the argument is *sound*, so we need to know whether all the premises are true.[20]

Premise 1

Habermas thinks that premise (1), the rules, norms or formal-pragmatic preconditions of discourse can be given a transcendental-pragmatic justification.[21] Roughly speaking, the idea is that whenever participants in discourse (who are as such always already oriented toward reaching consensus) assert *p*, they must, even if only counterfactually, assume that everyone ought to accept *p* as the result of an ideally prosecuted discourse. The ideality of discourse is preserved by certain implicit norms that are internal to the practice. Habermas does not give an exhaustive list of these implicit rules of discourse, but assumes that any exhaustive list will contain the following four rules:

(a) that nobody who could make a relevant contribution may be excluded;

(b) that all participants are afforded equal opportunities for participation;

(c) that participants must mean what they say;

(d) that communication must be free from internal and external compulsion, so that the yes/no stances that participants adopt towards critcizable validity-claims are motivated solely by the rational force of better reasons. (**OCCM**, p.356/**DEA**, p.62)[22]

Under (b) Habermas includes three further rules which are relevant to the derivation of (U) namely that:

(i) Everyone is allowed to question any assertion whatsoever;

(ii) Everyone is allowed to introduce any assertion whatsoever into the discourse;

(iii) Everyone is allowed to express his attitudes, desires and needs;

The 'transcendental-pragmatic' justification of premise (1) would take the form of a maieutic demonstration to the sceptical *participant* in discourse of the following points: (A) that he intuitively cannot but make assumptions concerning

rules of discourse; (B) that he can recognize them once they have been identified and described; and (G) that examples can corroborate the discourse ethicist's assertion that there are no alternatives to these assumptions (**MCCA**, p.97). The demonstration uses the device of performative self-contradiction to make these rules salient. A performative self-contradiction obtains when a participant implicitly, i.e. by virtue of the performative act of making an utterance, invokes rules which he explicitly, i.e. by the propositional content of his utterance, denies. For example, '*p*, but I do not mean *p*', or to use Moore's famous example, 'it is raining, but I do not believe it'. I am not going to discuss the propriety of this kind of 'transcendental-pragmatic' justification. I am content just to report Habermas's position that such a justification can be given for the above rules. Let us grant that there is a satisfactory transcendental-justification of premise (1), since nothing in my argument below will depend on it. Let us assume further that Habermas's proposed derivation does not depend on rules of discourse which are not contained in the above list from (a) to (d) so that his argument does not trade on hidden premises.

Premise 2

In his essay 'On the Cognitive Content of Morality' Habermas suggests that 'the normative conception of justification' that serves as the second premise is expressed in principle (D) (**OCCM**, p.355/**DEA**, p. 59). (D) states that:

> only those norms can claim validity that could meet with the acceptance of all concerned in practical discourse.(**OCCM**, p.354/**DEA**, p.59)[23]

Note that (D) specifies a necessary condition, namely that valid or impartially justified norms must be amenable to consensus in discourse. It does not state that

consensus is also a sufficient condition of validity. (D) does not state that what it is to be valid (or justified) is to be thus amenable to consensus. It leaves open the thought that there may be invalid or unjustified norms that are amenable to consensus in discourse. But why should we accept (D)?

(D) rests on the pragmatic theory of meaning. Until the late 1990s Habermas claims that an epistemic conception of truth and normative rightness are two specifications of a single underlying generic conception of validity. He contends, firstly, that meaning can be explicated by its validity basis - namely by participants' knowledge of the justifiability conditions of utterances; and secondly, that the validity (justifiability) of utterances connects necessarily with their amenability to consensus in ideally prosecuted discourse.[24] Habermas calls the consensus that would arise from an ideally prosecuted discourse a 'rationally motivated consensus'. We can formulate this idea of validity as follows:

For any *p*: if *p* is valid, then *p* is amen*able* rationally motivated consensus

Two clarifications are required here. First, the 'if then' is not a logical entailment but a pragmatic connection that inheres in our linguistic practices. As for the modal claim in the consequent, it refers to what it would be possible for real participants (not intelligible characters or super rational beings) to accept as a result of a real (not a hypothetical) discourse, but one which is ideally prosecuted in conformity with the above rules.

Now Habermas claims further that there is an analogy between truth and rightness. The analogy is explained by the fact that both of these values are specifications of the single underlying conception of validity.

For any utterance *p*: if *p* is true, then *p* is amenable to r.m.c.[25]

For any norm *n*: if *n* is right, then *n* is amenable to r.m.c.

Now it is easy to see that the latter specification of validity as rightness just is principle (D). We can check it against the formulation of (D) previously cited. In this respect the validity—consensus conditional contained in (D) provides a direct analogue in practical discourse with validity claims to truth in theoretical discourse.[26] Premise (2) rests on Habermas's pragmatic theory of meaning and on the analogy between truth and rightness.[27] That said, I think that (D) is, when suitably clarified, intuitively plausible. It amounts to the claim that, if a norm is justifiable, then it can be accepted by everyone in an ideally prosecuted discourse. We do not have to buy into the controversial underlying metaethical or linguistic theory, before we accept the principle. Let us, for the sake of argument, grant premise (2) as well.

Premise 3

Even if we accept premises (1) and (2) the proposed derivation fails, for premise (3) is clearly false. (U) states that:

> a norm is valid *if and only if* the foreseeable consequences and side effects of its general observance for the interests and value-orientations of *each individual* could be freely accepted *jointly* by *all* concerned. (**OCCM**, p.354)

(U) is a biconditional which states that the amenability to consensus on the basis of interests is a necessary and a sufficient condition of the validity of a norm. It is a *criterion* of validity in the strongest sense in which Wittgenstein used that term.[28] The amenability to consensus is not merely evidence for validity, it constitutes the validity of a norm. Formalising the argument can help to show what is wrong with it.[29] Let '*n*' be a variable ranging over all norms. Predicates 'V', 'C' and 'I' stand for 'is valid/justified,' 'is amenable to discursive consensus' and 'is amenable to a discursive consensus of interests' respectively.

(1)	R	premise
(2)	$\forall n$ (V$n \rightarrow$ Cn)	premise
(3)	(R & $\forall n$ (V$n \rightarrow$ Cn))	1,2, & I
(3*)	[(R & $\forall n$ (V$n \rightarrow$ Cn)) $\rightarrow$ $\forall n$ (V$n \leftrightarrow$I n)]	premise
(4)	$\forall n$ (V$n \leftrightarrow$I n)	3,3*, $\rightarrow$ E

The first problem is that it in no way follows from premises (1) and (2), the rules of discourse and principle (D), that the amenability to discursive consensus is a sufficient (as well as a necessary) condition of the validity of a norm. Nothing in the rules of discourse warrants this inference. According to the central plank of discourse meta-ethics - namely the alleged analogy between truth and rightness - (D) contains a necessary but not also a sufficient condition of normative validity. If it (D) were already a biconditional the situation would be worse. (D) would be the criterion of normative validity. The derivation of (U) would already contain the criterion of normativity - or something very close to it - in premise (2), which would itself then require derivation from non-moral premises.[30] Otherwise the programme of Discourse Ethics would be open to the charge of vicious circularity.

The second problem with the above argument lies in the difference between the indeterminate nature of the consensus, amenability to which is a necessary condition of validity according to (D), and the much richer notion of a consensus of interests, amenability to which is a necessary and sufficient condition of validity according to (U). (U) states that a norm is valid/justifiable, (i.e. there is sufficient reason to accept it) if and only if it satisfies or embodies

what Habermas calls a 'universalizable interest'. A further premise is needed that links the sufficiency of reasons with the existence of a universalizable interest in a norm's implementation (and all its foreseeable consequences). In the absence of such a premise, there is too big a gap between (D) and (U).

One possibility is that the inference to the richer notion of a consensus of interests in (U) is somehow warranted by the first premise. This may be where rule (b) (iii), 'Everyone is allowed to express his attitudes, desires and needs', comes in. The trouble with (b) (iii) is that, as Habermas himself concedes in the second edition of his 1983 essay, it is 'obviously irrelevant for theoretical discourses'. Despite this, Habermas adds curiously, '[i]t belongs to the pragmatic presuppositions of argumentation as such' (**MCCA**, p.89 n.72/**MKH**, p.99 n.71). But if the expression of individual desires and interests has nothing whatsoever to do with the search for truth, why should a rule permitting everyone to freely express their desires and interests count among the rules of discourse or argument in general? If, on the other hand, this rule figures in the premises as a precondition of *moral* argument or of *moral* discourse, then the suspicion of circularity is raised again.

William Rehg has suggested that the gap can be closed in the following way. Firstly, he elaborates meaning of 'norm' as a 'shared behavioural expectation' whose general observance resolves conflicts of action 'by regulating the satisfaction of the relevant interests of those involved (in light of a value or values the norm defines as having priority for all.'[31] Secondly, he adds a further premise, that participants in discourse find themselves 'in a modern pluralist society beset by conflicts of interest whose normative regulation can be convincingly based - should one decide for argued solutions at all - only on direct argumentation over which interest or value is to have priority in situations

of a given type'.[32] Rehg sees that 'grounding this assumption falls to a theory of modernity informed by a theory of communicative action.'[33] Technically speaking this is not just an *assumption* which can be discharged: it is a premise on which the argument for (U) rests. Once it is added, the proposed derivation of (U) depends in part on modernization theory.

However, for all Rehg's insightful and detailed elaboration of the hidden premises in the formal derivation of (U), he does not show how we can get from the necessary condition of validity in (D) to the biconditional in (U). He does not explain whence comes the requirement in (U) that the amenability to a consensus of interests be also a sufficient condition of validity. And that alone rules out Habermas's claim that a formal derivation of (U) from (1) and (2) can be provided.[34] This does not mean that there is no possible justification of (U). It just means that there is no apparent way to derive (U) from premises (1) and (2), and that such a derivation should therefore play no part in Habermas's justification programme.

IV. Modernization Theory and the Abduction of (U)

Habermas's work of the 1990s exhibits markedly less confidence that a formal derivation of (U) can be provided. Rather than attempt to provide a formal deduction of (U), Habermas is now content to make (U) plausible by adducing genealogical and historical arguments 'resting on assumptions of modernization theory' (**OCCM**, p.357/**DEA**, p.63). In *lieu* of a formal derivation, Habermas holds out the prospect of a weaker justification which does not depend on the logical derivation of (U). Habermas suggests that (U) follows from (1) and (2) as 'initially just an hypothesis won through abduction' (**DEA**, p.60/**OCCM**, p.354).

This justification strategy is weaker because (U) no longer follows by formal-logical entailment, but informally by 'abduction'.

'Abduction' is a term C. S. Peirce used to name the informal process by which inquirers come up with a best guess about which hypothesis to select and to subject to inductive testing. Peirce considered that the process of abduction is not just a matter of luck. It comprises a broad range of rational considerations such as, the purpose of the hypothesis, simplicity, elegance, explanatory scope, and compatibility with other beliefs.[35] In other words abduction is an inference to the best explanation in which a range of different pragmatic criteria fill out the relevant superlative. Habermas's claim now is that (U) suggests itself as the best explanation of the moral phenomenon in question, namely 'the "ought" character (*Sollgeltung*) of norms and the claims to validity raised in norm-related (or regulative) speech acts' (**MCCA**, p.44).

What makes (U) and the Discourse Theory of Morality the best explanation of the normativity of moral utterances? Habermas mentions two criteria, usefulness and intuitiveness. (U) must prove useful as a rule of moral argumentation in so far as it 'succeeds in selecting norms that are amenable to universal consensus'. Further it must not lead 'to counterintuitive results' (**OCCM**, p.355/**DEA**, p.60). The abductively won moral principle must be able to capture our intuitively most certain cases.

But are not the most common objections to principle (U) that it is not useful and deeply counterintuitive?[36] Of course it depends on what is meant by 'useful' and 'intuitive'. It may be that each norm that passes the stringent test of universalization contained in (U) is amenable universal consensus, even though very few do. Habermas's claim seems something like this: (U) is justified if it

passes some norms that are universally acceptable and no norms that are not and thus yields no result that is counterintuitive.

Now, just about the only valid norms embodying universalizable interests that Habermas adduces with any confidence are 'those that enshrine fundamental human rights' (**OCCM**, p.355/**DEA**, p.60). But it might be objected that, if only those norms enshrining fundamental human rights can be confidently expected to meet the condition that (U) imposes, then that itself is deeply counterintuitive. For there is a huge discrepancy in scope between everything we intuitively understand under the term 'immorality' e.g. lying, promise breaking, disloyalty, hypocrisy etc. and what intuitively falls under the concept of human rights violations.

Habermas has a good response to this objection. Moral intuitions are by their nature messy and the intuitive boundaries of morality vague. Any moral principle, whether or not it is based on the ideal of universal agreement, is bound to be selective. It is therefore reasonable to expect such a principle only to justify the central hard core of values and norms, the ones to which we are most deeply committed. It would be unreasonable to expect the norms that principle (U) selects to reflect the whole field of pre-reflective candidate values and norms. It is enough that (U) is intuitive in the first sense, that it validates some norms, however few, and that none of the norms it validates is counterintuitive. Further, Habermas claims, it is not a methodological consequence of his moral theory, that very few actual norms are capable of eliciting a universal consensus of interests. It is rather an effect of actual social and cultural change that the domain of morality has shrunk to a hard core of universally acceptable hence obligatory norms more or less coextensive with that of universal human rights (**JA**, p.91).

The following example may help to illustrate Habermas's point. Earlier last Summer there was agreement among a surprisingly wide political spectrum that the NATO intervention in Kosovo was *morally* justified. The *moral* justification of intervention was not the manifest insincerity and duplicity showed by the Serbian regime when negotiating the Rambrouillet accords, it was that intervention seemed to be the only viable means to halt the mass expulsions, ethnic cleansing, genocide, rape and torture being perpetrated by the Serbian militia in Kosovo. The moral norms the violation of which succeeded in uniting the international political community against the political and economic odds, were in fact none other than those enshrining universal human rights.

It might be thought that this example shows only that, given the paucity of recognized institutions and the lack of enforceable sanctions at the level of international law, human rights violations are sometimes sufficient to justify the engagement of the 'international community' in the internal political affairs of neighbouring countries, but that this shows nothing about the nature of moral norms in general. However, one has to recall that the primary justification of the military intervention was a *moral* and not a political one.[37] No other grounds would have sufficed. The example certainly demonstrates the peculiar ability of the norms governing human rights to elicit very widespread consensus. If Habermas is right that under modern conditions the standpoint of morality is indeed restricted to and preserved by those norms which can still successfully elicit universal agreement, then moral norms indeed become all the more important in their role of coordinating action. The Kosovo example shows this to be true. Thus it undermines the objection that Habermas's whole conception of morality is counterintuitive.

Moreover, the objection does not show that (U) fails to meet the criteria of usefulness and intuitiveness in the weaker sense specified above, and it is these conditions which the abductive justification of (U) brings to bear.

Nonetheless, Habermas's suggested abduction will only serve in *lieu* of a deduction, if the various criteria which it can show that (U) satisfies are jointly sufficient to justify it. His argument - indeed, if I am right it is the only viable argument for (U) - is roughly as follows. The abbreviations ***MT, PTM & MP*** indicate whether a particular consideration is based on modernization theory, Habermas's pragmatic theory of meaning or on moral phenomenology.

(1) If morality is a practice whose function is to regulate conflicts of interest between agents in the life-world (***MT & MP***); and

(2) if, under modern conditions, conflicts are settled by appeal to impartially justified norms (***MT & MP***);and

(3) if, there is no functional alternative to discourse/argumentation as a means of arriving at impartially justified norms and thereby resolving conflicts of interest in the life-world (***MT& MP***); and

(4) if the existence of the practice of discourse/argumentation in general presupposes idealizing rules of discourse, (a-d) which can be demonstrated to be necessary (i.e. reflexively ultimate) through the device of performative self-contradiction (***PTM***); and

(5a) if the meaning of the predicates right/wrong, and of moral utterances depends on their conditions of justifiability (***MT & PTM***); and

(5b) if the discursive justifiability of a norm can be elucidated by its necessary pragmatic connection with consensus in discourse as contained in principle (D) (***PTM***); and

(6) if interests provide reasons that justify norms (***MT & MP***); and

(7) if there exists a number of universalizable interests formed in the light of shared (or shareable) moral intuitions or values (***MT & MP***); and

(8) if it can be shown that there is a single rule of argumentation, (U), which is consistent with the rules of discourse and principle (D) (***PTM***); and

(9) if (U) can serve as a moral principle (regulating conflicts of interest) because,

   a. it selects some norms and each selected norm is one which every participant in discourse can accept in the light of their interests (***MP***); and

   b. it selects no norms which are inconsistent with our deepest moral intuitions (***MP***);

(10) then (U) is justified.

The long conjunction of considerations can be seen as the antecedents in a conditional of the form: if (a & b & c...), then U. This captures the hypothetical nature of the argument. Remember that this is not supposed to be a formally valid argument, but an informal abduction or inference to the best explanation that takes the place of the formal derivation of (U) in the original programme of Discourse Ethics. Writing out the argument schematically, as I have done here, makes clear that it rests very heavily on modernization theory. The antecedents (1), (2), (3), (5), (6), and (7) that license the inference to (U) as the best explanation, all rest in part or in whole on modernization theory.

It is true that Habermas has not entirely abandoned all hope that a formal derivation of (U), or one that is 'immanent' to the pragmatic theory of meaning, is possible. Indeed it is part of the programme of Discourse Ethics. For Habermas thinks that only a formal or 'immanent' derivation can completely allay the

sceptic's suspicion that rational reconstruction of morality rests on an ethnocentric fallacy (**DEA**, p.61). I have argued that a formal derivation of (U) from premises (1) and (2) is not possible. Now I want to claim that Habermas's recent concessions mean that a formal derivation of (U) is no longer even necessary. For in response to Rehg's criticisms, Habermas has conceded that principle (D), or premise (2), is itself partly based on modernization theory.

> If the practice of deliberation itself is regarded as the sole possible resource for a standpoint of impartial justification of moral questions, then the appeal to moral contents must be replaced by the self-referential appeal to the form of this practice. (D) expresses this understanding of the situation (**OCCM**, p. 353-4/**DEA**, p.59).

In other words, it falls to modernization theory to show that the antecedent holds, i.e. that, as moderns, we must seek impartial solutions to problems arising from conflicts of interests. For in the aftermath of the shattering of religious traditions and comprehensive metaphysical doctrines, there can be no further recourse to a universally shared set of substantial norms and values. Furthermore modernity theory must show that only the idealized procedure of moral discourse can provide a standpoint from which such conflicts can be resolved impartially. Because moral discourse contains within it a standard of impartial justification that any valid norm must meet, and thus can replace the example of a 'missing "transcendent good"' (**OCCM**, p.353/**DEA**, p.58). In this case, principle (D) and the alleged analogy between truth and rightness rest on modernization theory, not the pragmatic theory of meaning and the theory of communication. Consequently, even if (U) were to follow deductively from the conjunction of premises (1) and (2) Habermas's moral theory would still depend in part on modernization theory.

Habermas obviously thinks it much harder for the sceptic to reject 'the neutrality of discourse principle', (D), than the moral principle, (U). That is because (D) follows from the universality of the practice of argumentation, and from the fact that *for us* modern agents there is no alternative. To reject (D) would be thus to reject two sets of facts, reconstructive facts about the nature of discourse on the one hand, and historical and sociological facts about the role of discourse on the other. The question is, do these facts obtain, and can they be ascertained, independently of modernization theory. If, as I suspect, the answer to this question is 'no', then the sceptic will be able to claim that modernization theory [and thus also principle (D)] is also merely an *ethnocentric prejudice*. But if (D) is no more immune from sceptical suspicion than (U), there is no need for Habermas to continue to hold out the in my view forlorn hope that a logical derivation of (U) from the conjunction of (1) and (2) can be provided.

V. Conclusion

In section III I showed that there is no formal or 'immanent' derivation of principle (U), the central idea of Habermas's discourse theory of morality, from the premises of discourse theory alone - the rules of discourse and principle (D). In so far as it is justified, (U) remains just an 'abduction' or inference to the best explanation, and this inference leans heavily on modernization theory. This bears out my thesis that in the course of its development from the original programme the Discourse Theory of Morality has come to heavily on Habermas's theory of modernization.

What is the upshot of this development for the theory of modernization on the one hand and for the Discourse Theory of Morality on the other? Firstly, if my argument is correct, the Discourse Theory of Morality can no longer offer

the independent justificatory support to Habermas's modernization theory that the original programme of Discourse Ethics promised. The Discourse Theory of Morality rests very heavily on evidence provided by modernization theory. That the two theories are mutually consistent is therefore no surprise. It is only to be expected.

My claim is not that this fatally damages either Habermas's theory of modernization or his conception of modernity. According to modernization theory, there is a general, underlying pattern of socio-cultural development which has repercussions in each of the different value-spheres that separate out in the course of the development of modernity. Crudely speaking, the *Modernity Thesis* is supposed to capture that general phenomenon. If modernization really is a general phenomenon, it should leave its traces in the moral sphere. My point is simply that since Habermas's Discourse Theory of Morality rests on modernization theory, it cannot itself be offered as evidence for the truth of that theory. Not that modernization theory depends solely on that evidence. It depends also on a broad range of other considerations historical, legal, aesthetic, scientific etc.

Second, what is the upshot for the programme of Discourse Ethics as originally conceived, namely as a programme of justification of the moral standpoint? Put starkly I suggest that the Discourse Theory of Morality has not delivered the programme of philosophical *justification* if the moral standpoint that Discourse Ethics initially promised. It is, at best, a programme of the philosophical elucidation of moral normativity.

What is the difference between these two programmes? Well, conceived as a programme of the philosophical elucidation of moral normativity, the Discourse Theory of Morality can no longer take itself to be an answer to the

moral sceptic. The question of the moral sceptic is, 'Why be moral?' understood as a demand for justification. According to Habermas, the original programme of Discourse Ethics answers in the following way: by

(1) demonstrating to the sceptic through the device of performative contradiction that even he, to the extent that his utterances are meaningful and thus oriented towards reaching agreement, must implicitly recognize the pragmatic presupposition of argumentation;

(2) formulating the above as rules or implicit norms of discourse;

(3) showing that, to the extent that the sceptic can justify his utterances (which is assumed as a condition *sine qua non* of the ability to make meaningful utterances) he has always already recognized the principle of universalization (U). This is where the logical inference to (U) from premises (1) and (2) comes in. Of course, if must be assumed that the sceptic recognizes the laws of logic.

(4) challenging the sceptic to live a life without reliance or recourse to communicative action and discourse.

Having outlined this justification programme Habermas draws the following conclusion:

> If the sceptic has followed the argumentation that has gone on in his presence [(1-3) above *GF*] and has seen that his demonstrative exit from argumentation and action oriented toward reaching understanding leads to an existential dead end [(4) above *GF*], he may finally be ready to accept the justification of the moral principle that I have introduced. (**MCCA**, p.102)

However, we have seen that the Discourse Theory of Morality now concedes that both (D), and abduction of (U) rest heavily on modernization theory. And there is no reason the sceptic must accept that theory, even if he is committed, on pain of

performatively contradicting his own utterances, to recognize that the rules of discourse are binding on him too.

Thus to say that the Discourse Theory of Morality is a programme of elucidation is to admit that Habermas cannot avoid presupposing the truth of modernity theory as a premise in his only available argument for (U). Not only that, it is to admit that the programme does not rest on entirely non-moral premises. Both modernization theory and the abduction of (U) presuppose the existence of the moral standpoint. Modernization theory presupposes the existence of a standpoint of impartial justice as the historical bequest of religious tradition. The abduction of (U) presupposes the existence of some valid norms (namely those norms embodying universal human rights) against which the usefulness and intuitiveness of the moral principle (U) can be checked. Thus the justification of the moral standpoint proposed by the Discourse Theory of Morality presupposes exactly what the sceptic rejects.

Discourse theory still provides an answer to the question, 'Why be moral?', but in a much weaker sense than before. It aims to elucidate the self-understanding of agents who already recognize the normative meaning of moral utterances and the validity of moral norms; the self-understanding, that is, of modern moral agents.

All this by no mean implies that the outcome of the programme of Discourse Ethics is trivial. It is not obvious that there is a single principle of the validity of norms that is both useful and able to capture our deepest intuitions about what is morally right. Nor is it by any means obvious that, if there is one, it is (U). If Habermas can show that there is such a principle, and that it is (U), if he can demonstrate that his rational reconstruction of the pragmatic presuppositions

of communicative action and discourse is the best available hypothesis for making sense of our current moral practices, then he has accomplished a lot.

To see this one only need consider a very prominent, current line of argument against Habermas's conception of Discourse Ethics. On this popular but defeatist view Habermas should drop the moral-theoretical aspirations of Discourse Ethics, for in the end it fails as a theory of the validity of moral norms, and only succeeds as a theory of the democratic legitimacy of socio-political norms.[38] That such a judgment should be so widespread is testament enough that the Discourse Theory of Morality, understood as a programme of the philosophical elucidation of the moral standpoint, lacks nothing in controversy and ambition.

Abbreviations of Habermas's works referred to here are as follows: **BFN** = *Between Facts and Norms,* tr. W. Rehg (Cambridge: Polity Press, 1996) **CES** = *Communication and the Evolution of Society* (London: Heinemann, 1979): **DEA** = *Die Einbeziehung des Anderen*, (Frankfurt a/M: Suhrkamp, 1996). **DMUP =** *Die Moderne ein Unvollendetes Projekt* (Leipzig: Reklam, 1994): **ED =** *Erläuterungen zur Diskursethik* , (Frankfurt a/M: Suhrkamp, 1991) **FG** = *Faktizität und* Geltung (Frankfurt a/M: Suhrkamp, 1994) **JA =** *Justification and Application* (Cambridge: Polity Press, 1993); **MCCA** = *Moral Consciousness and Communicative Action* (Cambridge: Polity Press, 1990): **MKH =** Moralbewußtsein und kommunikatives Handeln (Frankfurt a/M: Suhrkamp, 1993): **OCCM** = 'On the Cognitive Content of Morality', *Proceedings of the Aristotelian Society,* 1997: **PDM =** *The Philosophical Discourse of Modernity* (Cambridge: Polity Presss, 1987): **PNK** = *Die Postnationale Konstellation: Politische Essays* (Frankfurt a/M: Suhrkamp, 1998): **RW** = 'Richtigkeit vs.

Wahrheit', *Deutsche Zeitschrift der Philosophie* 46 (1998) 2, pp. 179-208: **SE** = 'Sprechakttheoretischer Erläuterungen zum Begriff der kommunikativen Rationalität', in *Zeitschrift für Philosophische Forschung*, 50, 1996, pp. 65-91: **TCA 1** = *Theory of Communicative Action* (Boston: Beacon Press, vol. 1. 1984): **TCA 2 =** *Theory of Communicative Action* (Boston: Beacon Press, vol. 2 1987) **VE** = *Vorstudien und Ergänzungen zur Theorie des Kommunikativen Handelns* (Frankfurt a/M:, Suhrkamp, 1984).

[†] Significant parts of this paper are the fruit of a *Blockseminar* I co-taught last Summer at the University of Münster with Professor Josef Früchtl. Thanks to Josef and his students and to ERASMUS for funding the teaching exchange. Thanks also to Anthony Hatzimoysus, and my colleagues Steve Holland and Christian Piller.

[1] Henceforth I will call the original programme as set out in 1983 (**MKH: MCCA**) the programme of Discourse Ethics. I contrast this with the Discourse Theory of Morality, the shape of which which only clearly emerges in the mid 1990s even though Habermas coins the phrase as early as 1988. N.B. the excellent choice of the translators Lenhardt and Nicholsen to render the seminal 1983 essay, subtitled 'Notizen zu einer Begründungsprogramme' in English as 'Notes on a Programme of *Philosophical* Justification.' (my italics) A lot of potential confusion is sown because Habermas and his commentators do not signal the distinction between the meta-theoretical or meta-ethical justification that Discourse Ethics claims to provide, and the first-order justifiability of norms which, according to Habermas, constitutes the validity of moral norms; i.e the distinction between the *theoretical* justification that Discourse Ethics is, and the *moral* justifiability which it is about.

[2] This does not imply that the moral theory is justified on non-normative grounds. For the norms of discourse (see below) play a central role in Habermas's pragmatic theory of meaning.

[3] Originally Habermas called it a theory of 'practical discourse'. Only after 1988 does he develop the distinction between pragmatic, ethical and moral discourse. But retrospectively he realizes that (U) was all along a moral principle, at the centre of a moral theory. (**JA** p.vii)

[4] Discourse Ethics, claims Habermas, 'can be built into theories of the development of moral and legal consciousness at both the sociocultural and the ontogenetic levels and in this way can be made susceptible to indirect corroboration' (**MCCA**, p.98).

[5] Not that Habermas has abandoned the original programme entirely; he still holds out the hope that a formal derivation can be provided (**DEA**, p.61), although he now concedes, in face of the

detailed and persuasive arguments of William Rehg, that one premise of the proposed deduction - principle (D) - rests on modernization theory. (See IV below)

[6] (**TCA1**, p.157-216)

[7] See Hegel's *Elements of the Philosophy of Right*, §124 Remark: 'The right of the subject's particularity to find satisfaction, or ... the right of subjective freedom, is the pivotal and focal point in the difference between *antiquity* and *modernity*.' Also §273 Addition (Hotho and Gans) 'The principle of the modern world in general is freedom of subjectivity, according to which all essential aspects present in the social totality develop and enter into their right.' (Cambridge: Cambridge University Press, 1991) Habermas makes the same point in terms of developmental psychology. Modern subjects develop post-conventional cognitive competencies and form abstract ego identities: that is, they are capapable of acting on the basis of self-chosen principles rather than traditional values or external authority, and their sense of self is uncoupled from traditional roles, practices and values etc. of a particular form of life (**CES**, p.69-95, **TCA2**, p.92-107, **PNK**, p. 221-231).

[8] (**TCA1** 157-215)

[9] In this Habermas sides with Talcott Parsons against Weber's theory of rationalization (**TCA2**, p. 283-99).

[10] Here Habermas argues as a neo-Weberian along with Horkheimer, Adorno and the later Marcuse *contra* Parsons.

[11] It is because Habermas sees the advent of modernity as an opportunity for achieving social stability and legitimacy, whilst widening the scope for individual autonomy, that he resists the trend of some postmodernist writers to say good-bye and good-riddance to the project of modernity and its opportunities. See for example his 1980 Adorno Prize lecture 'Modernity - an Unfinished Project' (**DMUP** p.33-55). This refusal to throw the baby out with the bathwater of modernization lies at the heart of his polemic against some forms postmodernism inspired by

Nietzsche and Heidegger in (**PDM**).

[12] Habermas now claims that this task falls both to moral and to ethical discourses. However, in the early 1980s he did not make this distinction and tended to treat practical and moral discourse as equivalent.

[13] 'Even if those who conceived themselves as "moderns" always invented an idealized past to imitate, none the less this modernity now conscious of itself has to justify this choice of model with its own standards and to create everything normative out of itself. Modernity has to stabilize itself on the basis of the only authority that its has left standing, namely on the basis of reason.' (**PNK**, p.198) See also (**PNK**, p.196) on modern Romanticism.

[14] See also (**CES**, pp.69-95 & **PMT**, pp.149-205).

[15] 'Denn der Diskurs kann, dank seiner normative gehaltvollen Kommunikationsvoraussetzungen, jene Beschränkungen, die der Rechtfertigungspraxis mit dem Entwurf eines moralischen Universums auferlegt werden, *aus sich selbst heraus* erzeugen.'

[16] More recently he formulates premise (1) as 'the implicit content of the universal preconditions of argumentation' (**OCCM**, p.355/**DEA**, p.61).

[17] Habermas also claims that premise (2) consists in the participants' knowledge of 'what it means to discuss hypothetically whether norms of action should be adopted' (**MCCA**, pp. 92 & 198).

[18] Otherwise we could derive (U) like this: $(a \,\&\, \neg a) \rightarrow (U)$.

[19] I am using 'true' here in the very broad sense of whatever designated epistemic value is passed from premises to conclusion by valid inference.

[20] See note 19.

[21] Note that it is a common mistake to think that the whole argument can be given (and thus that the conclusion - the 'truth' of principle (U) can be established by) a transcendental-pragmatic justification. This is not and never was Habermas's position.

[22] See W. Rehg, *Insight and Solidarity* (Berkeley: University of California Press, 1994), p. 62ff.

[23] See also (**MCCA**, p.66: **FG**, p.138/**BFN**, p.107).

[24] Habermas (**JA**, p.53) 1990: 'With the assertoric meaning of his utterance, the speaker raises a criticizable claim to the validity of the asserted proposition; and since we have no direct access to uninterpreted conditions of validity, "validity" [*Gültigkeit*] must be understood epistemically as "validity [*Geltung*] that is established for us'. Habermas (**DEA**, p.53-4/**OCCM**, p.350-1) 1997: 'What interests me... is the possibility of understanding the concept of truth, cleansed of all correspondence connotations, as a special case of validity. [...] The rightness of moral norms (or normative utterances) ...can then be understood in analogy with the truth of assertoric sentences.'

[25] Habermas treats this claim as equivalent with the following: For any normative utterance *q*: if *q* is right, then *q* is amenable to rationally motivated consensus.

[26] 'Discourse ethics, then, stands or falls with two assumptions: (a) that normative claims to validity have cognitive meaning and can be treated *like* claims to truth and (b) that the justification of norms and commands requires that a real discourse be carried out and thus cannot occur in a strictly monological form...' (**MCCA**, p.68).

[27] Most surprisingly, in his most recent essay, Habermas abandons his long-held view that truth is epistemic and concedes that truth, unlike normative rightness, is a justification-transcendent concept.

> 'Truth' is a justification-transcendent concept, which cannot even be captured with the concept of ideally justified assertability. It points rather to truth conditions, which to a certain extent have to be fulfilled by reality itself. By contrast the meaning of 'rightness' [*Richtigkeit* i.e. correctness G.F.] can be reduced to ideally justified acceptability. (**RW**, p.188)

As a consequence he must give up the claim that truth and rightness are specifications of a single generic conception of validity. For once we allow that truth - call it Truth^{jt} - outstrips justification, we allow the possibility that there are unjustifiable Truths^{jt}. But why should we expect anyone in discourse to accept such a Truth^{jt} ? Conversely there may be justifiable Falsehoods^{jt} which

everyone has reason to accept. Hence there can be no necessary connection between Truths^{jt} and the amenability to rationally motivated consensus. In this way Habermas's recent abandonment of an epistemic (or justification-immanent) concept of truth in favour of a much richer, non-epistemic or justification-transcendent concept derails the analogy between truth and rightness in respect of their pragmatic connections with consensus, and blocks the claim that truth and rightness are specification of a single underlying conception of validity.

[28] *The Blue and Brown Books* (Oxford: Blackwell, 1958), p. 25, and *Wittgenstein's Lectures on the Foundations of Mathematics*, *Cambridge 1939* (Sussex: Harvester Press, 1976), p. 164, cited in *Wittgenstein: Meaning and Mind*, P. M. S. Hacker (Oxford: Blackwell 1990), p. 558. See also Rogers Albritton, 'On Wittgenstein's Use of the Term "Criterion"', in *Journal of Philosophy* LVI No. 22 1959, pp. 845-857.

[29] Nothing in my argument depends on the formalization given here. But it shows at a glance, and in high resolution, what is wrong with the derivation. It is important to write the argument out formally in order to test Habermas's claim that a logical derivation of (U) is possible.

[30] In 1986 Seyla Benhabib pointed out that premise (2) 'depending on how it is interpreted...reads as if it were simply equivalent to some version of U'. In her view premise (2) already contains an implicit reference to a consensus based on common interests. She concludes that the derivation of (U) is viciously circular, and that (U) is either redundant or rests on hidden normative premises. *Critique, Norm and Utopia* (New York: Columbia University Press, 1986) pp. 307-8. I think this is a little uncharitable. Subsequently Habermas has made it clear that (D) is not a biconditional, and that it does not already make reference to an interest-based consensus. The problem remains that there is a lacuna between principle (D) and the much stronger principle (U).

[31] W. Rehg, 'Discourse and the Moral Point of View: Deriving a Dialogical Principle of Universalization', *Inquiry* 34 (1991), p. 36.

[32] *Ibid.* p. 38.

[33] *Ibid.* p. 38

[34] Albrecht Wellmer claims that Habermas's derivation of principle (U) is self-evidently false. 'Ethics and Dialogue' in *The Persistence of Modernity* (Cambridge: Polity Press, 1991), p. 182. He supposes that it is obvious what Habermas's argument is. I take the view that it is worth showing what the argument is, and that it is unsound because premise (3) is not true.

[35] Christopher Hookway, *Peirce* (London: Routledge, 1985), pp.223-8

[36] For the objection that (U) is redundant see Thomas McCarthy 'Practical Discourse: On the Relation of Morality to Politics' in *Ideals and Illusions,* (Cambridge MA: MIT Press, 1991), p. 198; and 'Legitimacy and Diversity Dialectical Reflections on Analytical Distinctions', *Cardozo Law Review* 17/4-5 (1996), pp.1083-1127. See also Maeve Cooke, 'Habermas and Consensus', *European Journal of Philosophy* 1:3, 1993 pp.257-8; and *Language and Reason: A Study of Habermas's Pragmatics* (Cambridge MA: MIT Press, 1994), pp.153-4. For the objection that it is counterintuitive see Seyla Benhabib who complains that Habermas's distinction between moral questions (of justice) and ethical questions of what is good for me/us 'contradicts our deepest moral intuitions'. 'How can Kohlberg and Habermas defend a position, which so totality contradicts our intuitions and the phenomenology of our moral experience?' S. Benhabib, *Selbst im Kontext: Kommunikative Ethik im Spannungsfeld von Feminismus, Kommunitarismus und Post-Moderne* (Suhrkamp, Frankfurt a/M, 1995), pp. 200-1.

[37] I am leaving aside the moral question of whether it was morally justified to intervene in the way NATO did, merely bombarding Serb targets from the air.

[38] See on this point Albrecht Wellmer, 'Ethics and Dialogue', in *The Persistence of Modernity*, (Cambridge: Polity Press, 1991); Agnes Heller, 'The Discourse Ethics of Habermas: Critique and Appraisal', *Thesis Eleven* 10/11, (1984-5), pp.5-17; and Simone Chambers, *Reasonable Democracy: Jürgen Habermas and the Politics of Discourse* (Ithaca; New York: Cornell University Press, 1996), p.145, 'The U-principle does not make sense as a criterion of moral truth,

but it does make sense as a criterion of democratic legitimacy'.